

日 本 国 特 許 庁

JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office

出 願 年 月 日

Date of Application:

2001年 7月 4日

出 願 番 号

Application Number:

特願2001-202918

出 願 人

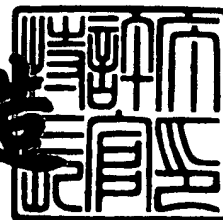
applicant(s):

株式会社日立製作所

2001年 8月17日

特 許 庁 長 官
Commissioner,
Japan Patent Office

及 川 耕 造



CERTIFIED COPY OF
PRIORITY DOCUMENT

出証番号 出証特2001-307258

【書類名】 特許願

【整理番号】 K01000481A

【あて先】 特許庁長官殿

【国際特許分類】 G06F 3/06

【発明者】

【住所又は居所】 神奈川県小田原市中里 3 2 2 番地 2 号 株式会社日立製作所 RAIDシステム事業部内

【氏名】 吉田 晃

【特許出願人】

【識別番号】 000005108

【氏名又は名称】 株式会社日立製作所

【代理人】

【識別番号】 100075096

【弁理士】

【氏名又は名称】 作田 康夫

【手数料の表示】

【予納台帳番号】 013088

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 ディスクアレイ制御装置における共有メモリコピー機能

【特許請求の範囲】

【請求項 1】

ホストコンピュータとのインターフェースを有するチャネルインターフェース部と、磁気ディスク装置とのインターフェースを有するディスクインターフェース部と、前記磁気ディスク装置に対しリード／ライトされるデータを一時的に格納するキャッシュメモリ部と、前記チャネルインターフェース部および前記ディスクインターフェース部と前記キャッシュメモリ部との間のデータ転送に関する制御情報および前記磁気ディスク装置の管理情報を格納する共有メモリ部と、

前記チャネルインターフェース部および前記ディスクインターフェース部と前記キャッシュメモリ部を接続する手段と、前記チャネルインターフェース部および前記ディスクインターフェース部と前記共有メモリ部を接続する手段と、前記各部を駆動する電源供給手段を有し、前記ホストコンピュータからのデータのリード／ライト要求に対し、前記チャネルインターフェース部は、

前記ホストコンピュータとのインターフェースと前記キャッシュメモリ部との間のデータ転送を実行し、前記ディスクインターフェース部は、前記磁気ディスク装置と前記キャッシュメモリ部との間のデータ転送を実行することにより、データのリード／ライトを行うディスクアレイ制御ユニットを、複数ユニット有するディスクアレイ制御装置であって、前記複数のディスクアレイ制御ユニット内の前記共有メモリ部間を接続する手段と、前記複数のディスクアレイ制御ユニット内の前記キャッシュメモリ部間を接続する手段を有し、前記ディスクアレイ制御ユニット内の前記共有メモリ部と他の前記ディスクアレイ制御ユニット内の前記共有メモリ部間で前記共有メモリ部格納データのコピー処理が可能であることを特徴とするディスクアレイ制御装置。

【請求項 2】

ホストコンピュータとのインターフェースを有するチャネルインターフェース部と、磁気ディスク装置とのインターフェースを有するディスクインターフェース部と、前記磁気ディスク装置に対しリード／ライトされるデータを一時的に格

納するキャッシュメモリ部と、前記チャンネルインターフェース部および前記ディスクインターフェース部と前記キャッシュメモリ部との間のデータ転送に関する制御情報および前記磁気ディスク装置の管理情報を格納する共有メモリ部と、

前記チャンネルインターフェース部および前記ディスクインターフェース部と前記キャッシュメモリ部を接続する手段と、前記チャンネルインターフェース部および前記ディスクインターフェース部と前記共有メモリ部を接続する手段と、前記各部を駆動する電源供給手段を有し、前記ホストコンピュータからのデータのリード／ライト要求に対し、前記チャンネルインターフェース部は、前記ホストコンピュータとのインターフェースと前記キャッシュメモリ部との間のデータ転送を実行し、

前記ディスクインターフェース部は、前記磁気ディスク装置と前記キャッシュメモリ部との間のデータ転送を実行することにより、データのリード／ライトを行うディスクアレイ制御ユニットを、複数ユニット有するディスクアレイ制御装置であって、前記複数のディスクアレイ制御ユニット内の前記共有メモリ部間を接続する手段と、前記複数のディスクアレイ制御ユニット内の前記キャッシュメモリ部間を接続する手段を有し、該接続手段を介して、前記ディスクアレイ制御ユニット内の前記共有メモリ部と他の前記ディスクアレイ制御ユニット内の前記共有メモリ部間で前記共有メモリ部格納データのコピー処理が可能であり、コピー処理中においても、ディスクアレイ制御ユニット内の前記チャンネルインターフェース部および前記ディスクインターフェース部から、前記コピー処理の領域に対する前記共有メモリ部格納データのリード／ライト処理可能であることを特徴とするディスクアレイ制御装置。

【請求項 3】

前記複数のディスクアレイ制御ユニット内の前記複数のチャンネルインターフェース部および前記複数のディスクインターフェース部と前記複数のキャッシュメモリ部との間は、前記複数のディスクアレイ制御ユニット間に跨るスイッチを用いた相互結合網によって接続され、前記複数のチャンネルインターフェース部および前記複数のディスクインターフェース部と前記複数の共有メモリ部との間は、前記複数のディスクアレイ制御ユニット間に跨るスイッチを用いた相互結合網に

よって接続され、前記ディスクアレイ制御ユニット内の前記共有メモリ部と他の前記ディスクアレイ制御ユニット内の前記共有メモリ部間で前記共有メモリ部格納データのコピー処理が可能であることを特徴とするディスクアレイ制御装置。

【請求項 4】

前記ディスクアレイ制御ユニット内の前記共有メモリ部の格納データを他前記ディスクアレイ制御ユニット内の前記共有メモリ部に前記相互結合網を介して転送する手段を持ち、前記転送手段により格納データのどの領域まで転送が終了しているかを記録する手段を持つことを特徴とする請求項 1 または 2 に記載のディスクアレイ制御装置。

【請求項 5】

前記ディスクアレイ制御ユニット内の前記共有メモリ部の格納データを他前記ディスクアレイ制御ユニット内の前記共有メモリ部に前記相互結合網を介して転送する手段を持ち、前記転送手段は前記相互結合網の使用頻度をモニタする機能を持ち、前記相互結合網の使用頻度が低い時に前記転送を行う機能を持つことを特徴とする請求項 1 または 2 に記載のディスクアレイ制御装置。

【請求項 6】

前記ディスクアレイ制御ユニット内の前記共有メモリ部の格納データを他前記ディスクアレイ制御ユニット内の前記共有メモリ部に前記相互結合網を介して転送する手段を持ち、前記転送中に転送先の前記共有メモリ部に読み出し要求があった場合、転送終了領域であるか否かを判断し、転送終了領域であれば前記転送先共有メモリ部から格納データを読み出し、転送終了未の領域であれば前記転送元共有メモリ部から格納データを読み出し、

前記転送中に転送先の前記共有メモリ部に書き込み要求があった場合、前記転送先共有メモリ部および前記転送元共有メモリ部の対応領域に格納データを書き込む機能を持つことにより、前記転送中の他処理に影響を及ぼす事無く、非同期に転送を行う事が出来る機能を持つ事を特徴とする請求項 1 または 2 に記載のディスクアレイ制御装置。

【請求項 7】

前記ディスクアレイ制御ユニット内の前記共有メモリ部の格納データを他前記

ディスクアレイ制御ユニット内の前記共有メモリ部に前記相互結合網を介して転送する手段を持ち、前記転送中にアプリケーションプログラムは転送処理中であることを意識せずにリード／ライト処理可能である事を特徴とする請求項 1 または 2 に記載のディスクアレイ制御装置。

【請求項 8】

前記ディスクアレイ制御ユニット内の前記共有メモリ部内に通常動作に使用しないリザーブ領域を有し、前記ディスクアレイ制御ユニット内の前記共有メモリ部内の格納データを他前記ディスクアレイ制御ユニット内の前記共有メモリ部内のリザーブ領域に前記相互結合網を介して転送する手段を持つことを特徴とする請求項 1 または 2 に記載のディスクアレイ制御装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、ディスクアレイ制御装置に関し、特に、データを複数の磁気ディスク装置に格納するディスクアレイ装置の制御装置技術に関する。また、複数のサーバおよびパソコンを接続した装置に関しても、本発明は利用可能である。

【0002】

【従来の技術】

今般のコンピュータシステムにおいて、処理性能の向上に対する期待は大きく、特にディスクサブシステムの I/O 性能向上に対する要求は高い。磁気ディスクを記憶媒体とするディスクサブシステム（以下「サブシステム」という。）の I/O 性能は半導体記憶装置を記憶媒体とするコンピュータの主記憶の I/O 性能に比べて、3～4 桁程度小さく、従来からこの差を縮めること、すなわちサブシステムの I/O 性能を向上させる努力がなされている。

【0003】

また、銀行、証券、電話会社等に代表される大企業では、従来各所に分散していたコンピュータおよびストレージを、データセンターの中に集中化してコンピュータシステムおよびストレージシステム構成することにより、コンピュータシステムおよびストレージシステムの運用、保守、管理に要する費用を削減する傾

向にあり、特に大型／ハイエンドのストレージシステムには、数百台以上のホストコンピュータへ接続するためのチャネルインターフェースのサポート（コネクティビティ）、数百テラバイト以上の記憶容量のサポートが要求されている。

【0004】

一方、近年のオープン市場の拡大、今後予想されるストレージ・エリア・ネットワーク（SAN）の普及により、大型／ハイエンドのストレージシステムと同様の高機能・高信頼性を備えた小規模構成（小型筐体）のストレージシステムへの要求が高まっている。

【0005】

サブシステムのI/O性能を向上させるための1つの方法として、複数の磁気ディスク装置でサブシステムを構成し、データを複数の磁気ディスク装置に格納する、いわゆるディスクアレイと呼ばれるシステムが知られている。ディスクアレイの場合、上位コンピュータからのI/Oを記録する複数の磁気ディスク装置と、上位コンピュータのI/Oを受け複数の磁気ディスク装置へ転送するディスクアレイ制御装置から構成されるのが一般的である。この中で、大規模接続・大容量への要求に対しては、従来の大型／ハイエンドのディスクアレイ制御装置を複数接続して超大規模なディスクアレイ制御装置を構成する方法が考えられる。ディスクアレイ制御装置にはディスクアレイ制御装置に関する制御情報（例えば、ディスクアレイ制御装置内のキャッシュメモリの管理情報等）を格納する共有メモリを保持することが知られている。

【0006】

複数のディスクアレイ制御装置を接続することにより、共有メモリが複数のディスクアレイ制御装置に分散することになる。共有メモリが複数のディスクアレイ制御装置に分散することにより共有メモリ領域のコピーに伴う転送処理性能が、従来の一つのディスクアレイ制御装置によりシステムが構築されていた場合に比べ、ディスクアレイシステムの性能に与える影響が大きくなる。

【0007】

例えば、従来技術では、図2に示すようにホストコンピュータ50とディスクアレイ制御装置2との間のデータ転送を実行する複数のチャネルIF部11と、

磁気ディスク装置 5 とディスクアレイ制御装置 2 間のデータ転送を実行する複数のディスク I/F 部 1 2 と、磁気ディスク装置 5 のデータを一時的に格納するキャッシュメモリ部 1 4 と、ディスクアレイ制御装置 2 に関する制御情報（例えば、チャンネル I/F 部 1 1 およびディスク I/F 部 1 2 とキャッシュメモリ部 1 4 との間のデータ転送制御に関する情報、磁気ディスク装置 5 に格納するデータの管理情報）を格納する共有メモリ部 1 3 とを備え、1 つのディスクアレイ制御装置 2 内において、共有メモリ部 1 3 およびキャッシュメモリ部 1 4 は全てのチャンネル I/F 部 1 1 およびディスク I/F 部 1 2 からアクセス可能な構成となっている。

【0008】

このディスクアレイ制御装置 2 では、チャンネル I/F 部 1 1 およびディスク I/F 部 1 2 と共有メモリ部 1 3 との間、およびチャンネル I/F 部 1 1 およびディスク I/F 部 1 2 とキャッシュメモリ部 1 4 との間は、相互結合網 2 3、および相互結合網 2 4 でそれぞれ接続される。

【0009】

チャンネル I/F 部 1 1 は、ホストコンピュータ 50 と接続するためのインターフェースおよびホストコンピュータ 50 に対する入出力を制御するマイクロプロセッサ（図示せず）を有している。また、ディスク I/F 部 1 2 は、磁気ディスク装置 5 と接続するためのインターフェースおよび磁気ディスク装置 5 に対する入出力を制御するマイクロプロセッサ（図示せず）を有している。また、ディスク I/F 部 1 2 は、RAID 機能の実行も行う。

【0010】

この従来のディスクアレイ制御装置 2 では、ディスクアレイ制御装置内に共有メモリ部 1 3 が存在する為、共有メモリ間で共有する情報を持つ必要が無く、共有メモリ間でコピーが必要になった場合についても共有メモリ部が同一装置内にある為、転送に伴う相互接続網の競合等の他アクセスに対する影響が小さい。

【0011】

また、コピー機能について、米国特許 5,680,640 に開示されている従来技術では、旧記憶装置から新記憶装置へとデータを移行する際にオンラインで実行する方法が示されている。ここでは、新記憶装置内に、旧記憶装置内ボリュ

ームの各アドレス（トラック）毎にテーブルを有し、旧記憶装置から新記憶装置へデータ移行が完了したかどうかをトラック毎に記憶する。移行中にホストからのI/O要求があった場合には、そのテーブルを参照して動作を決定する。例えばリード要求があった場合、テーブルを参照して、そのリード要求のあったレコード（ブロック）が新記憶装置に移行されているかどうかをチェックし、旧記憶装置からのデータが移行されていない時には旧記憶装置からデータを読み込む。

【 0 0 1 2 】

新記憶装置内にデータがあれば新記憶装置内からデータを読み出す。また、書き込み要求があった場合には、新記憶装置にデータ書き込みを行い、テーブルの更新を行う。本方式は、記憶装置を置き換える場合のコピー機能について示されているが、本方式を複数のディスクアレイ制御装置により構成されるシステムに応用することは可能である。

【 0 0 1 3 】

【発明が解決しようとする課題】

米国特許 5, 6 8 0, 6 4 0 に開示されている従来技術では、コピー元の記憶装置内全てのラックについてテーブルを保持しなければならない。通常テーブルには半導体メモリが使用され

コピー元とコピー先各々にテーブルを持つことにより、コストが高くなる。

【 0 0 1 4 】

また、この方法では、データ移行中のホストからの書き込みでコピー先にのみデータを書き込む為、移行中にどちらかの記憶装置に障害が発生した場合、どちらの記憶装置もデータに矛盾のある状態となってしまう。また、この方式は、記憶装置の置き換え時のデータコピーについての技術であり、動作中の複数のディスクアレイ制御装置間のデータコピーについて記述したものではない。

【 0 0 1 5 】

【課題を解決するための手段】

上記課題およびディスクアレイ制御装置間の共有メモリ間コピー機能を実現するため、本発明におけるディスクアレイ制御装置は、下記の構成を取る。目的は、ホストコンピュータとのインターフェースを有するチャネルインターフェース

部と、磁気ディスク装置とのインターフェースを有するディスクインターフェース部と、磁気ディスク装置に対しリード／ライトされるデータを一時的に格納するキャッシュメモリ部と、チャンネルインターフェース部およびディスクインターフェース部とキャッシュメモリ部との間のデータ転送に関する制御情報および磁気ディスク装置の管理情報を格納する共有メモリ部と、チャンネルインターフェース部およびディスクインターフェース部とキャッシュメモリ部を接続する手段と、チャンネルインターフェース部およびディスクインターフェース部と共有メモリ部を接続する手段と、各部を駆動する電源供給手段を有し、

ホストコンピュータからのデータのリード／ライト要求に対し、チャンネルインターフェース部は、ホストコンピュータとのインターフェースとキャッシュメモリ部との間のデータ転送を実行し、ディスクインターフェース部は、磁気ディスク装置とキャッシュメモリ部との間のデータ転送を実行することにより、データのリード／ライトを行うディスクアレイ制御ユニットを、複数ユニット有するディスクアレイ制御装置であって、複数のディスクアレイ制御ユニット内の共有メモリ部間を接続する手段と、複数のディスクアレイ制御ユニット内のキャッシュメモリ部間を接続する手段を有し、共有メモリ部間を接続する手段とキャッシュメモリ部間を接続する手段は、

おのおの独立に動作し、相互に影響を与えない手段を有し、ディスクアレイ制御ユニット内のチャンネルインターフェース部およびディスクインターフェース部から、他のディスクアレイ制御ユニット内の前記共有メモリ部のデータ、またはキャッシュメモリ部のデータをリード／ライト可能であり、ディスクアレイ制御ユニット内の共有メモリ部と他のディスクアレイ制御ユニット内の共有メモリ部間で格納データのコピー処理が可能であることを特徴とするディスクアレイ制御装置により達成される。

【0016】

好ましくは、複数のディスクアレイ制御ユニット内の複数のチャンネルインターフェース部および複数のディスクインターフェース部と複数のキャッシュメモリ部との間には、複数のディスクアレイ制御ユニット間に跨るスイッチを用いた相互結合網によって接続し、複数のチャンネルインターフェース部および複数のディス

クインターフェース部と複数の共有メモリ部との間は、複数のディスクアレイ制御ユニット間に跨るスイッチを用いた相互結合網によって接続する。

【 0 0 1 7 】

共有メモリの内容を、他ディスクアレイ制御ユニット内の他共有メモリにコピーする際に、複数のディスクアレイ制御ユニット間に跨るスイッチを用いた相互結合網を介してデータが移行される。

【 0 0 1 8 】

コピーに先立ち、コピー開始アドレスとコピー終了アドレスおよびコピー開始情報をコピー先共有メモリ部および共有メモリ間に跨るスイッチ部の情報レジスタに登録を行う。このコピー開始情報の登録により、コピー元のマイクロプログラムは、コピーが終了したものとし、

次の処理に移行可能となる。ハードは、コピー開始情報の登録完了により、以降のコピー先共有メモリへの読み出しアクセス発生時、コピー情報レジスタとアクセスアドレスの比較により該アドレスがコピー完了未領域か否かの情報を得、コピー完了未領域へのアクセスの場合は、

コピー元共有メモリへのアクセスに切り替えコピー元共有メモリから該当データを読み出しリクエスト発行元にデータを送出する。また、リクエストがコピー完了済み領域へのアクセスであった場合コピー先共有メモリからデータを読み出し、リクエスト発行元にデータを送出する。リクエスト発行元のマイクロプログラムは、該当エリアがコピー中であることを意識する必要は無い。

【 0 0 1 9 】

その他、本願が開示する課題、およびその解決方法は、発明の実施形態の欄および図面により明らかにする。

【 0 0 2 0 】

【発明の実施形態】

以下、本発明の実施例を図面を用いて説明する。

〔実施例 1〕

図 1 および図 3 に、本発明の一実施例を示す。

【 0 0 2 1 】

図1に示すように、ディスクアレイ制御装置1は複数のディスクアレイ制御ユニット1-2から構成される。ディスクアレイ制御ユニット1-2は、ホストコンピュータ50とのインターフェース部（チャンネルIF部）11と、磁気ディスク装置5とのインターフェース部（ディスクIF部）12と、共有メモリ部13と、キャッシュメモリ部14を有し、チャンネルIF部11およびディスクIF部12と共有メモリ部13の間は、ディスクアレイ制御ユニット1-2内において直接に接続されている。また、複数のディスクアレイ制御ユニット1-2の間では、共有メモリ部13は相互結合網22を介して接続され、キャッシュメモリ部14は相互結合網21を介して接続されている。

【0022】

すなわち、相互結合網21、あるいは相互結合網22を介して、全てのチャンネルIF部11およびディスクIF部12から、全ての共有メモリ部13、あるいは全てのキャッシュメモリ部14へアクセス可能な構成となっている。共有メモリ部13内には、コピー開始情報および転送情報を格納するコピー情報レジスタ15と共有メモリ16を持つ。

【0023】

図1にて、ディスクアレイ制御ユニット1-2間の共有メモリ領域のコピー処理について考えると、チャンネルIF部11内のマイクロプログラムにより共有メモリ領域のコピー処理の起動をかける為に、チャンネルIF部11よりコピー元の共有メモリ領域17が存在する共有メモリ部13内のコピー情報レジスタ15にコピー元共有メモリアドレスおよびコピー先共有メモリアドレスとコピー開始を示すバリッドビットをセットする。同様に、コピー先の共有メモリ領域18が存在する他ディスクアレイ制御ユニット1-2内の共有メモリ部13内のコピー情報レジスタ15にもコピー情報をセットする。

【0024】

コピー情報レジスタ15へのコピー情報登録完了をコピー先共有メモリ部13からコピー要求元のチャンネルIF部のマイクロプログラムに知らせることにより、マイクロプログラムはコピー処理完了とし他処理に移行可能となる。コピー元共有メモリ領域17が存在する共有メモリ部13は、コピー情報レジスタのコピ

ー実行ビットがバリッドとなったことにより、共有メモリ16内のコピー元領域17の値を読み出し、相互結合網24を介し、他ディスクアレイ制御ユニット1-2内の共有メモリ部13に転送する。転送データ到着により、コピー先共有メモリ領域18を保持する他ディスクアレイ制御ユニット1-2内共有メモリ部13は、転送データをコピー先領域18に書き込む。

【0025】

同様な処理をコピー領域全てについて行い、全ての領域のコピー終了により、各々の共有メモリ部13内のコピー情報レジスタ15内のコピー実行ビットのバリッドをOFFにする。コピー実行中を示すコピー情報レジスタ15内のコピー実行ビットがバリッドの場合、コピー先共有メモリ部13内のコピー先領域18へのコピー処理に伴うアクセス以外のアクセスがあった場合、コピー先共有メモリ部13内の共有メモリ16にアクセスするリクエストは全て、コピー情報レジスタ15内のコピー終了未領域アドレスとアドレス比較することにより、アクセス領域がコピー終了未領域であった場合は、該当アクセスが読み出し要求の場合、ハードによりコピー元共有メモリ部13にリクエストを転送し、コピー元領域17からデータを読み出し、アクセス要求元にデータ転送を行う。

【0026】

該当アクセスが、書込みアクセスの場合は、コピー先領域18の該当アドレス部とコピー元領域17の該当アドレス部に書込みを行う。アクセス領域がコピー終了領域およびコピー対象領域以外の場合は、コピー先共有メモリ16の該当領域に対し、読み出しおよび書込み動作を行う。

【0027】

図3によりディスクアレイ制御ユニット1-2間の共有メモリ領域のコピー処理におけるコピー先の共有メモリ部13内の処理について考えると、マイクロプログラムにより共有メモリ領域のコピー処理の起動をかける為に、ディスクアレイ制御ユニット1-2間を跨る相互結合網22を介し、共有メモリアクセスパス200により、コピー先の共有メモリ領域18が存在する共有メモリ部13内のコピー情報レジスタ15内のコピー元開始アドレス102、コピー先開始アドレス103およびコピー先終了アドレス104にコピー情報をセットする。更に、

コピー先の共有メモリ領域18が存在する共有メモリ部13内のコピー情報レジスタ15内のコピー実行ビット101にコピー開始を示すバリッドビットをセットする。

【0028】

コピー実行ビット101がセットされていることにより、コピー情報レジスタ15内のアドレス生成論理107、コピー領域判定論理108およびコピー終了判定論理109が機能することになる。コピー情報レジスタ15へのコピー情報登録完了を共有メモリ部13からコピー要求元のチャンネルIF部のマイクロプログラムに知らせることにより、マイクロプログラムはコピー処理完了とし他処理に移行可能となる。

【0029】

もちろん、コピー先共有メモリ16へのコピー領域全ての書込み終了を待って、他処理に移行する方式としても良い。コピー先共有メモリ部13は、コピー実行ビット101がセットされていることにより、コピー処理に伴うコピーリクエストに対し、共有メモリ内のコピー先領域18に対し、コピー情報を書き込む。コピー処理に伴うコピーリクエスト受付により、コピー情報レジスタ15内のカウンタ105を更新し、コピー先開始アドレス103と加算することにより、コピー実行アドレス106に現在のコピー実行アドレスをセットする。

【0030】

コピーリクエストを受け付けるたびに、カウンタ105をカウントアップし、コピー実行アドレス106を更新する。コピー先終了アドレス104とコピー実行アドレス106の値をコピー終了判定論理109により比較することにより、コピー終了を判定する。コピー終了判定により、コピー実行ビット101をリセットし、コピー処理終了となる。コピー元共有メモリ部にも同様なコピー終了判定論理を持つことにより、コピー元共有メモリ部にコピー終了を知らせる必要はなくなる。

【0031】

もちろん、コピー終了をコピー元共有メモリ部およびマイクロプログラムに知らせる方式としても良い。次に、コピー処理中の、コピー処理以外の共有メモリ

16へのリクエストに対する処理について考えると、コピー処理中のコピー処理以外の共有メモリ16へのリクエストに対しては、共有メモリ部13内でコピー処理以外の共有メモリ16へのリクエストを優先とした方が性能的に有利である。コピー処理以外の共有メモリ16へのリクエスト受付により、共有メモリ部13は、コピー情報レジスタ15内のコピー実行ビット101がセットされている場合、コピー実行中と判断し、コピー領域判定論理108により、コピー実行アドレス106とコピー先終了アドレス104の値とリクエストアドレスの値を比較することにより、リクエストアドレスがコピー終了未領域へのアクセスか否かを判定する。コピー領域判定論理108により、リクエストアドレスがコピー終了未領域へのアクセスと判定した場合、アドレス生成論理107によりコピー元開始アドレス102の値とコピー先開始アドレス103の差分およびリクエストアドレスとの加算により、コピー元共有メモリへのアクセスアドレスを求め、コピー元共有メモリ部に他共有メモリアクセスパス202により、相互結合網22を介し、リクエストを転送する。

【0032】

書込みリクエストの場合、コピー終了未領域への書込みでも、自共有メモリ16の該当アドレスにも書き込む方式を取った方が、障害および性能面で有利である。コピー終了未領域への書込みの場合、更新済である情報を共有メモリ部13内に持つことにより（図示せず）、以降の該当領域のリクエストに対しても自共有メモリ16内へのリクエストとして受け付けることが可能となり、コピー元領域を含む他ディスクアレイ制御ユニット内のコピー元共有メモリ部にリクエストを転送する必要がなくなる。

【0033】

コピーまた、コピー領域判定論理108により、リクエストアドレスがコピー終了領域およびコピー対象領域以外へのアクセスと判定した場合、アドレス生成論理107によりリクエストアドレスを選択し、自共有メモリアクセスパス201を介し、自共有メモリ16内にアクセスする。

〔実施例2〕

図4に、本発明の他の実施例を示す。

図4に示すように、複数のディスクアレイ制御ユニット1-2からなるディスクアレイ制御装置1の構成は、複数のディスクアレイ制御ユニット1-2を跨ぐ、チャンネルIF部11およびディスクIF部12とキャッシュメモリ部14間の接続構成および共有メモリ部13間の接続構成を除いて、実施例1の図1に示す構成と同様である。図4に示すように、ディスクアレイ制御ユニット1-2内のキャッシュメモリ部14と他ディスクアレイ制御ユニット1-2内のキャッシュメモリ部14は、筐体間キャッシュスイッチ140を介して、筐体間キャッシュメモリバス141により相互に接続されている。

【0034】

障害を考え筐体間キャッシュスイッチ140は、冗長性を持たせ二重化構造となっている。また、同様に筐体間キャッシュメモリバス141についても二重化構成となっている。筐体間キャッシュメモリバス141および筐体間キャッシュスイッチ140を介することにより、複数のディスクアレイ制御ユニット1-2間で他ディスクアレイ制御ユニット1-2内のキャッシュメモリ部14にアクセスすることが可能となる。

【0035】

一般にキャッシュおよびディスクへアクセスするバスのデータ転送は、数キロバイトの転送が発生する為大規模であるのに対し、共有メモリをアクセスするバスのデータ転送は、データが制御情報の為、数バイト程度で数サイクルの規模である。相互結合網が一つの場合、共有メモリをアクセスするバスとキャッシュおよびディスクへアクセスするバスが同じ相互結合網を使用することになり、小規模アクセスである共有メモリアクセスの性能低下が生じることになる。このため、図4に示すように、ディスクアレイ制御ユニット1-2内の共有メモリ部13と他ディスクアレイ制御ユニット1-2内共有メモリ部13は、筐体間キャッシュメモリスイッチ140とは別の筐体間共有メモリスイッチ130を介して、筐体間共有メモリバス131により相互に接続されている。

【0036】

障害を考え筐体間共有メモリスイッチ130は、冗長性を持たせ二重化構造となっている。また、同様に筐体間共有メモリバス131についても二重化構成と

なっている。筐体間キャッシュメモリスイッチ140および筐体間キャッシュメモリバス141と筐体間共有メモリスイッチ130および筐体間共有メモリバス131は、各々独立に動作可能である。共有メモリ部13内のコピー情報レジスタ15内の構造は、図3にて示してあるコピー情報レジスタ15の構造と同一である。

【0037】

また、図4に示すように、筐体間共有メモリスイッチ130内にも同様にコピー情報レジスタ15を持つ構造となっている。図4によりディスクアレイ制御ユニット1-2間の共有メモリ領域のコピー処理における筐体間共有メモリスイッチ130内の処理について考えると、マイクロプログラムにより共有メモリ領域のコピー処理の起動をかける為に、ディスクアレイ制御ユニット1-2間を跨る筐体間共有メモリバス131および筐体間共有メモリスイッチ130を介し、コピー先の共有メモリ領域18が存在する共有メモリ部13内のコピー情報レジスタ15内のコピー元開始アドレス、コピー先開始アドレスおよびコピー先終了アドレスにコピー情報をセットする。

【0038】

更に、コピー先の共有メモリ領域18が存在する共有メモリ部13内のコピー情報レジスタ15内のコピー実行ビットにコピー開始を示すバリッドビットをセットする。同様に、筐体間共有メモリスイッチ130内のコピー情報レジスタ15内のコピー元開始アドレス102、コピー先開始アドレス103およびコピー先終了アドレス104にコピー情報をセットする。更に、筐体間共有メモリスイッチ130内のコピー情報レジスタ15内のコピー実行ビット101にコピー開始を示すバリッドビットをセットする。コピー実行ビット101がセットされていることにより、コピー情報レジスタ15内のアドレス生成論理107、コピー領域判定論理108およびコピー終了判定論理109が機能することになる。

【0039】

コピー先の共有メモリ領域18が存在する共有メモリ部13内および筐体間共有メモリスイッチ130内のコピー情報レジスタ15へのコピー情報登録完了をコピー先の共有メモリ領域18が存在する共有メモリ部13内および筐体間共有

メモリスイッチ130からコピー要求元のチャンネルIF部のマイクロプログラムに知らせることにより、マイクロプログラムはコピー処理完了とし他処理に移行可能となる。

【0040】

もちろん、コピー先共有メモリ16へのコピー領域全ての書込み終了を待って、他処理に移行する方式としても良い。コピー先共有メモリ部13は、コピー実行ビットがセットされていることにより、コピー処理に伴うコピーリクエストに対し、共有メモリ内のコピー先領域18に対し、コピー情報を書き込む。筐体間共有メモリスイッチ130は、コピー処理に伴うコピーリクエスト受付により、コピー情報レジスタ15内のカウンタ105を更新し、コピー先開始アドレス103と加算することにより、コピー実行アドレス106に現在のコピー実行アドレスをセットする。

【0041】

コピーリクエストを受け付けるたびに、カウンタ105をカウントアップし、コピー実行アドレス106を更新する。コピー先終了アドレス104とコピー実行アドレス106の値をコピー終了判定論理109により比較することにより、コピー終了を判定する。コピー終了判定により、コピー実行ビット101をリセットし、コピー処理終了となる。コピー元共有メモリ部にも同様なコピー終了判定論理を持つことにより、コピー元共有メモリ部にコピー終了を知らせる必要はなくなる。もちろん、コピー終了を筐体間共有メモリスイッチ130およびコピー先共有メモリ部13より、コピー元共有メモリ部およびマイクロプログラムに知らせる方式としても良い。次に、コピー処理中の、コピー処理以外の共有メモリ16へのリクエストに対する処理について考えると、コピー処理中のコピー処理以外の共有メモリ16へのリクエストに対しては、筐体間共有メモリスイッチ130内でコピー処理以外の共有メモリ16へのリクエストを優先とした方が性能的に有利である。

【0042】

コピー処理以外の共有メモリ16へのリクエスト受付により、筐体間共有メモリスイッチ130は、コピー情報レジスタ15内のコピー実行ビット101がセ

ットされている場合、コピー実行中と判断し、コピー領域判定論理108により、コピー実行アドレス106とコピー先終了アドレス104の値とリクエストアドレスの値を比較することにより、リクエストアドレスがコピー終了未領域へのアクセスか否かを判定する。コピー領域判定論理108により、リクエストアドレスがコピー終了未領域へのアクセスと判定した場合、アドレス生成論理107によりコピー元開始アドレス102の値とコピー先開始アドレス103の差分およびリクエストアドレスとの加算により、コピー元共有メモリへのアクセスアドレスを求め、コピー元共有メモリ部13に筐体間共有メモリバス131により、リクエストを転送する。

【0043】

また、コピー領域判定論理108により、リクエストアドレスがコピー終了領域およびコピー対象領域以外へのアクセスと判定した場合、アドレス生成論理107によりリクエストアドレスを選択し、筐体間共有メモリバス131を介し、該当共有メモリ領域を含むディスクアレイ制御ユニット1-2内の共有メモリ部13にアクセスする。筐体間共有メモリスイッチ130にコピー情報レジスタ15が無い場合には、コピー先領域18へのコピー処理以外のリクエストが発生した場合に、アクセス元のディスクアレイ制御ユニット1-2から筐体間共有メモリバス131および筐体間共有メモリスイッチ130を介し、コピー先領域18を含む他ディスクアレイ制御ユニット1-2内の共有メモリ部13にアクセスした後、コピー未領域と判定され、再び筐体間共有メモリバス131および筐体間共有メモリスイッチ130を介して、コピー元領域17を含む他ディスクアレイ制御ユニット1-2内の共有メモリ部13にアクセスする必要があるのに対し、筐体間共有メモリスイッチ130にコピー情報レジスタ15を持つことにより、コピー先領域18を含む他ディスクアレイ制御ユニット1-2内の共有メモリ部13にアクセスせずにアクセス領域がコピー未領域か否かの判定を得ることができ、処理の高速化および筐体間共有メモリバス131の専有時間を削減でき性能面で有利である。

【0044】

本実施例によれば、共有メモリ部間でコピー処理を行う場合、コピー実行元の

マイクロプログラムは、コピー情報をコピー情報レジスタ15に登録することにより、コピー処理終了となり、次処理に移行することが可能となる。コピー情報レジスタ15へのコピー情報登録以降は、コピー中であることを意識せずに全ての共有メモリへのアクセスが可能となる。コピー対象領域へのアクセスについても、アドレスの変更およびリクエストの振り分けをマイクロプログラムは意識する必要が無い。

【0045】

また、コピー処理動作より通常処理を優先とする論理を組み込むことにより、コピー処理に伴う、他処理への性能面の影響を少なくし、また、キャッシュアクセスの筐体間接続網と共有メモリの筐体間接続網を別々に持つことにより、共有メモリ部のコピー処理により、キャッシュアクセスに影響を及ぼすことがなくなる。また、共有メモリ部間の相互結合網を介しての転送手段において相互結合網の使用頻度をモニタする機能を持たせ、相互結合網の使用頻度が低い時に共有メモリ部格納データの転送を行うことによりサブシステム性能の低下を抑制可能である。

〔実施例3〕

図5に、本発明の他の実施例を示す。

図5に示すように、共有メモリ部13内の共有メモリ16内の構成を除いて、実施例1の図1に示す構成と同様である。図5に示すように、各々のディスクアレイ制御ユニット1-2内の共有メモリ部13内の共有メモリ16内にリザーブ領域19を持つことにより、ディスクアレイ制御ユニット1-2間の共有メモリ16の再構成および再配置が可能となる。チャンネルIF部11およびディスクIF部12とキャッシュメモリ14との間のデータ転送に関する制御情報および磁気ディスク装置5の管理情報を格納する共有メモリ部13は、各々自ディスクアレイ制御ユニット1-2内の情報を保持するほうが性能的に有利である。共有メモリ部13が複数のディスクアレイ制御ユニット1-2に分散されて実装された場合、自ディスクアレイ制御ユニット1-2内の情報を常に自ディスクアレイ制御ユニット1-2内の共有メモリ部13に保持することは難しいと考えられる。

【0046】

図5により、動作中の共有メモリ再配置について考えると、モニタリング機構等（図示せず）により、自ディスクアレイ制御ユニット1-2内の情報が他ディスクアレイ制御ユニット1-2内の共有メモリ部13に格納されていることを検出した場合、マイクロプログラムは自ディスクアレイ制御ユニット1-2内の情報が格納されている他ディスクアレイ制御ユニット1-2内の共有メモリ部13内のコピー情報レジスタ15にコピー元共有メモリアドレスおよびリザーブ領域アドレスとコピー開始を示すバリッドビットをセットする。

【0047】

同様に、自ディスクアレイ制御ユニット1-2内の共有メモリ部13内のコピー情報レジスタ15にコピー元共有メモリアドレスおよびリザーブ領域アドレスとコピー開始を示すバリッドビットをセットする。コピー情報レジスタ15へのコピー情報登録完了をコピー元共有メモリ部13から再配置要求元のチャンネルIF部のマイクロプログラムに知らせることにより、マイクロプログラムは再配置処理完了とし他処理に移行可能となる。コピー元領域17が存在する共有メモリ部13は、コピー情報レジスタのコピー実行ビットがバリッドとなったことにより、共有メモリ16内のコピー元領域17の値を読み出し、相互結合網24を介し、他ディスクアレイ制御ユニット1-2内の再配置要求元の共有メモリ部13に転送する。転送データ到着により、再配置要求元共有メモリ部13は、転送データをリザーブ領域19に書き込む。同様な処理をコピー領域全てについて行い、全ての領域のコピー終了により、各々の共有メモリ部13内のコピー情報レジスタ15内のコピー実行ビットのバリッドをOFFにする。

【0048】

コピー実行中を示すコピー情報レジスタ15内のコピー実行ビットがバリッドの場合、コピー元共有メモリ部13内のコピー元領域17へのコピー処理に伴うアクセス以外のアクセスがあった場合、コピー元共有メモリ部13内の共有メモリ16にアクセスするリクエストは全て、コピー情報レジスタ15内のコピー終了未領域アドレスとアドレス比較することにより、アクセス領域がコピー終了未領域およびコピー対象領域以外の場合は、コピー元共有メモリ部13内のコピー元共有メモリ16の該当領域に対し、読み出しおよび書込み動作を行う。アクセ

ス領域がコピー終了領域の場合は、該当アクセスが読み出し要求の場合、コピー元共有メモリ部 1 3 内のコピー元共有メモリ 1 6 の該当領域からデータを読み出し、アクセス要求元にデータ転送を行い、該当アクセスが、書込みアクセスの場合は、コピー元領域 1 7 の該当アドレス部と再配置要求元の共有メモリ部 1 3 内のリザーブ領域 1 9 の該当アドレス部に書込みを行う。

【 0 0 4 9 】

本実施例によれば、共有メモリ部間で再配置を行う場合、再配置要求元のマイクロプログラムは、コピー情報をコピー情報レジスタ 1 5 に登録することにより、再配置処理終了となり、次処理に移行することが可能となる。コピー情報レジスタ 1 5 へのコピー情報登録以降は、コピー中であることを意識せずに全ての共有メモリへのアクセスが可能となる。コピー対象領域へのアクセスについても、アドレスの変更およびリクエストの振り分けをマイクロプログラムは意識する必要が無い。また、コピー処理動作より通常処理を優先とする論理を組み込むことにより、コピー処理に伴う、他処理への性能面の影響を少なくし、また、キャッシュアクセスの筐体間接続網と共有メモリの筐体間接続網を別々に持つことにより、共有メモリ部のコピー処理により、キャッシュアクセスに影響を及ぼすことがなくなる。また、共有メモリ部間で、メモリ部格納データの転送中にアプリケーションプログラムは転送処理中であることを意識せずにリード／ライト処理可能である。

【 0 0 5 0 】

【発明の効果】

本発明によれば、複数台のディスクアレイ制御装置を 1 つのディスクアレイ制御装置として運用しようとする場合、複数のディスクアレイ制御装置に共有メモリが分散されて搭載された場合の、共有メモリ間でのコピー処理による性能低下を抑え、台数に比例した性能を出せるディスクアレイシステムを提供すること、また、ディスクアレイ制御装置が有する機能を、性能低下を抑えて複数台のディスクアレイ制御装置で実現することが可能となる。

【図面の簡単な説明】

【図 1】

本発明によるディスクアレイ制御装置の構成を示す図。

【図 2】

図 1 に示すディスクアレイ制御ユニット内の詳細構成を示す図。。

【図 3】

従来のディスクアレイ制御装置の他の構成を示す図。

【図 4】

本発明によるディスクアレイ制御装置の他の構成を示す図。

【図 5】

本発明によるディスクアレイ制御装置の他の構成を示す図。

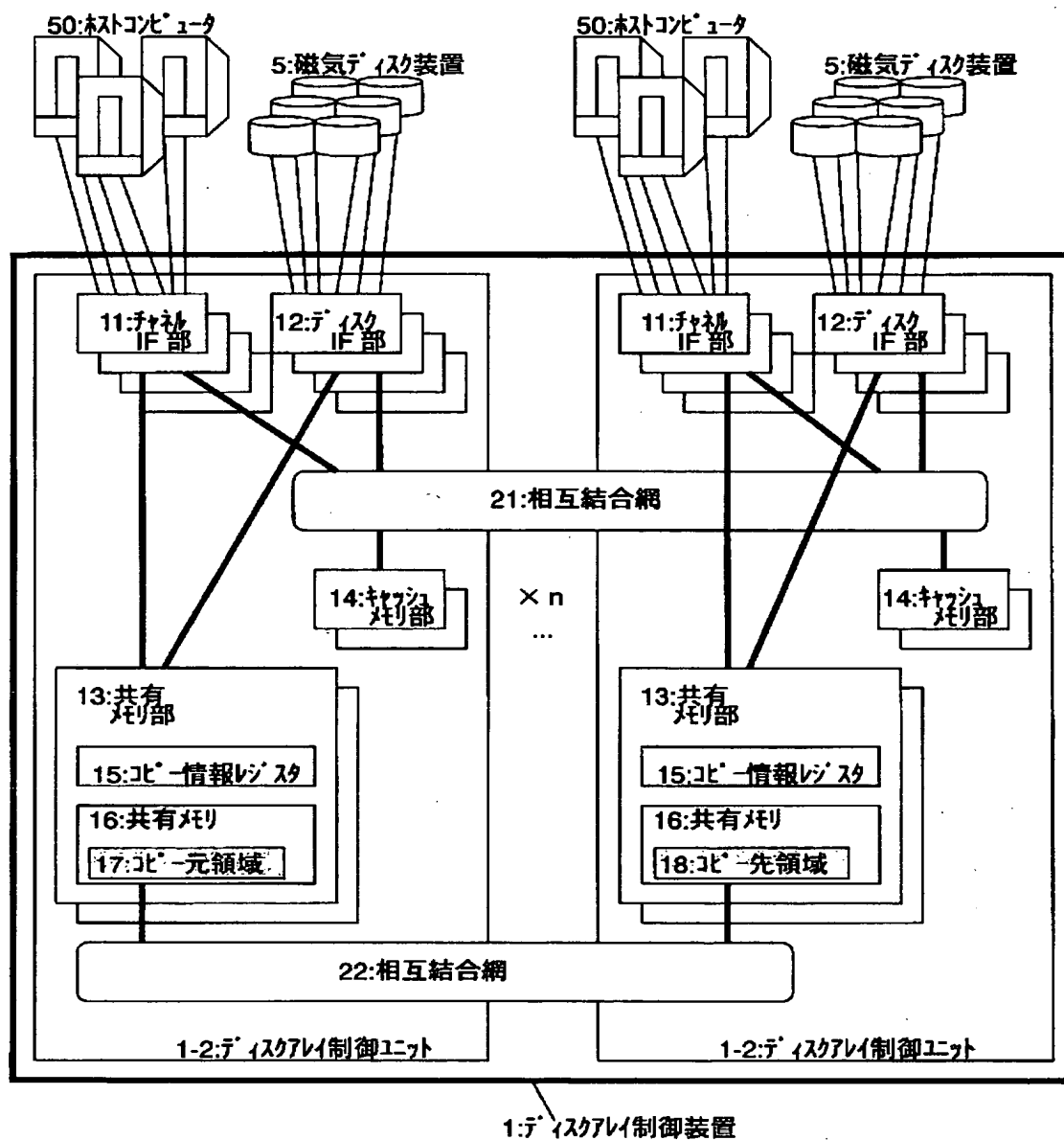
【符号の説明】

1, 2 : ディスクアレイ制御装置, 1 - 2 ... ディスクアレイ制御ユニット、5
... 磁気ディスク装置, 1 1 ... チャンネル I F 部、1 2 ... ディスク I F 部、1 3 ... 共
有メモリ部、1 4 ... キャッシュメモリ部、1 5 ... コピー情報レジスタ、2 1, 2
2, 2 3, 2 4 ... 相互結合網、5 0 ... ホストコンピュータ

【書類名】 図面

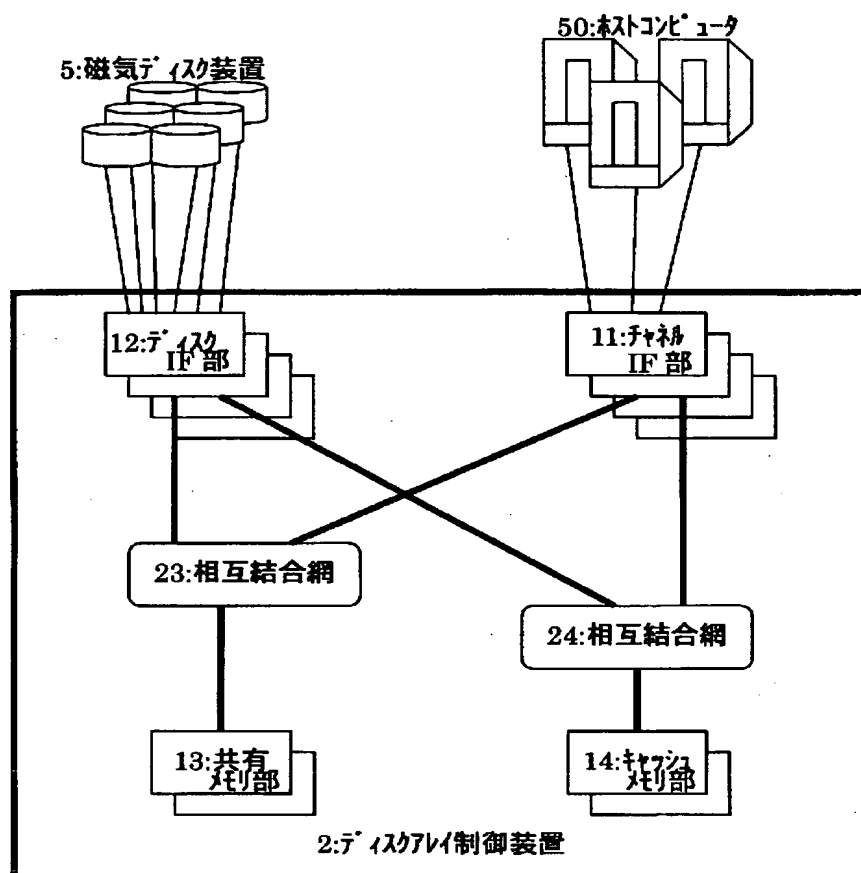
【図 1】

図 1



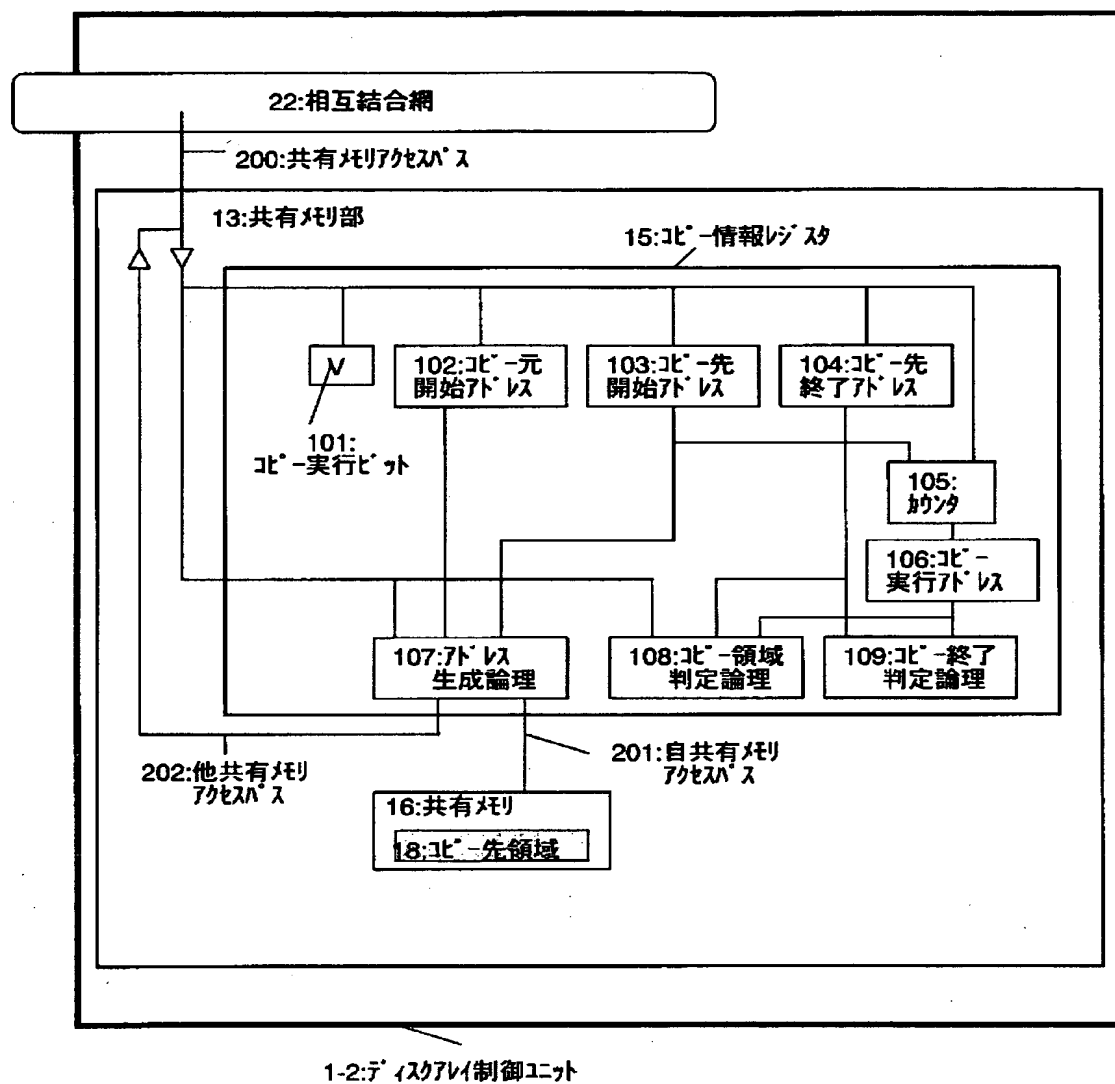
【図 2】

図 2



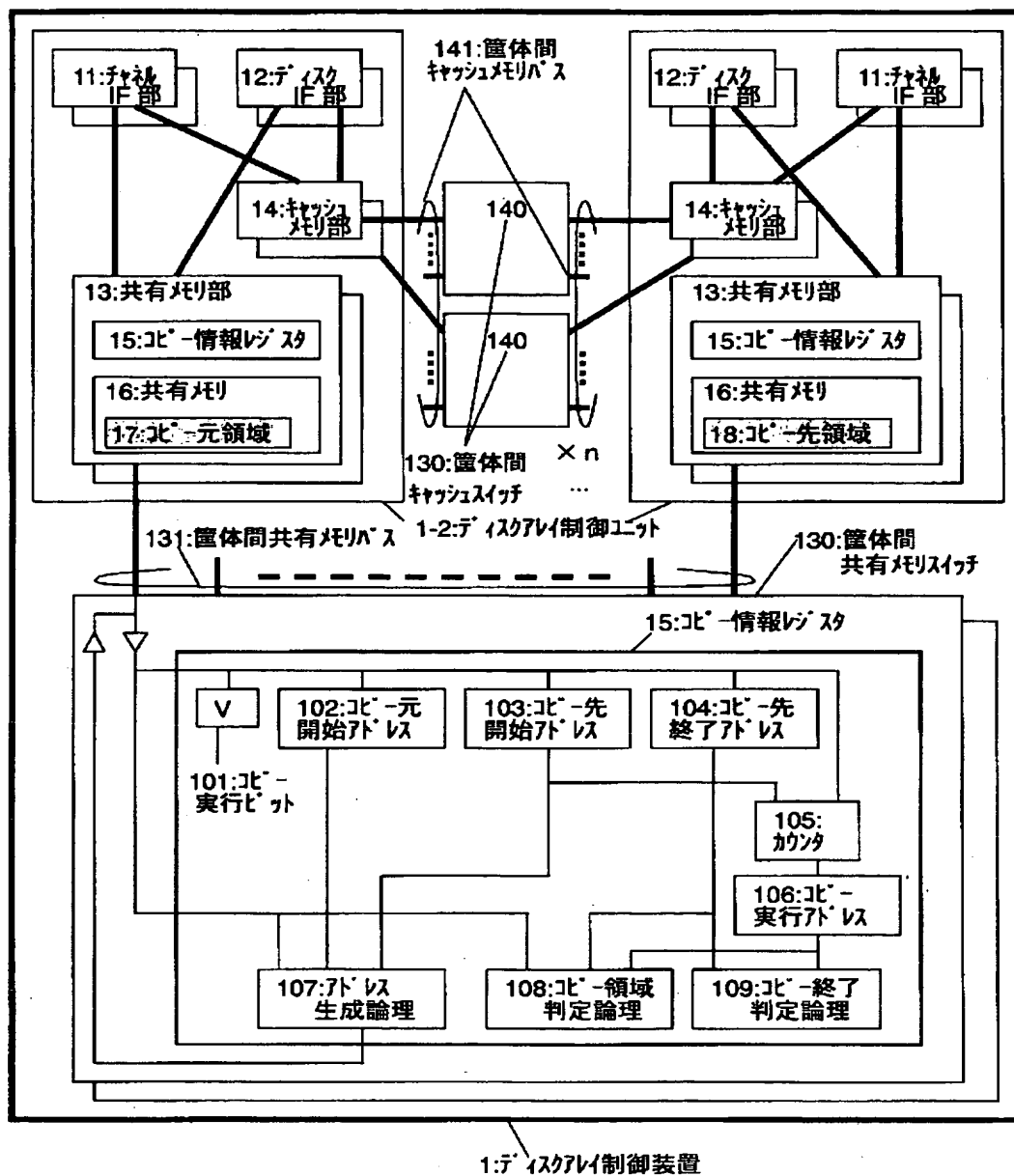
【図 3】

図 3



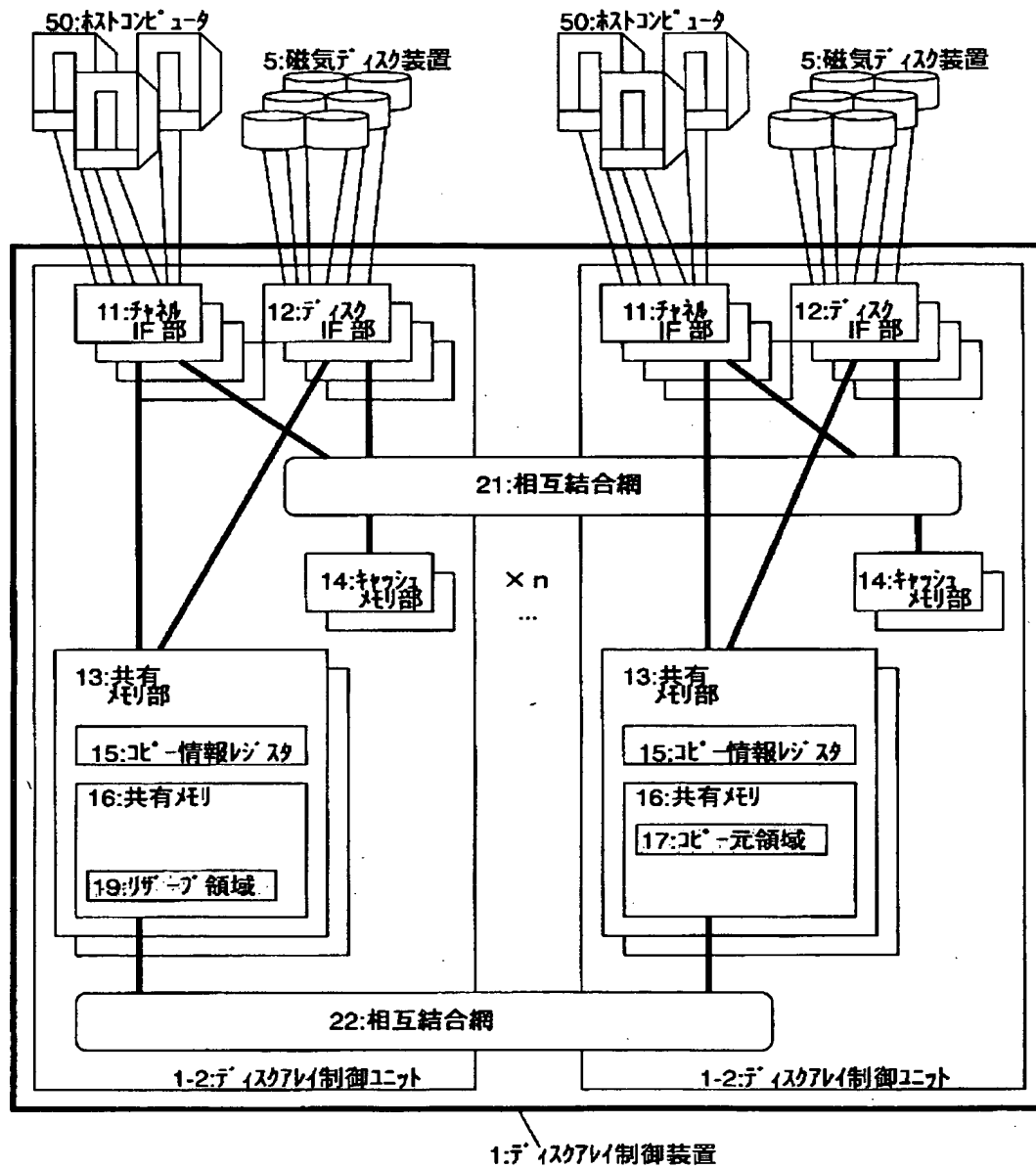
【図 4】

図 4



【図 5】

図 5



【書類名】 要約書

【要約】

【課題】

複数台のディスクアレイ制御装置を1つのディスクアレイ制御装置として運用できるようにし、複数のディスクアレイ制御装置間での共有メモリ部のコピー処理による性能低下を抑え、台数に比例した性能を出せるディスクアレイシステムを提供することにある。

【解決手段】

上記課題は、チャンネル I F 部と、ディスク I F 部と、キャッシュメモリ部と、共有メモリ部と、チャンネル I F 部およびディスク I F 部とキャッシュメモリ部を接続する手段と、チャンネル I F 部およびディスク I F 部と共有メモリ部を接続する手段と、前記各部を駆動する電源供給手段を有し、データのリード／ライトを行うディスクアレイ制御ユニットを、複数ユニット有するディスクアレイ制御装置であって、複数のディスクアレイ制御ユニット内の共有メモリ部間を接続する手段と、複数のディスクアレイ制御ユニット内のキャッシュメモリ部間を接続する手段を有し、ディスクアレイ制御ユニット内にコピー情報レジスタを有し、また、複数のディスクアレイ制御ユニット内の共有メモリ部間を接続する共有メモリスイッチ部にコピー情報レジスタを有することを特徴とするディスクアレイ制御装置により達成される。

【効果】

複数のディスクアレイ制御ユニット間でのコピー処理の性能の向上が可能で、ディスクアレイ制御ユニットの台数に比例してディスクアレイ制御装置の性能を向上できる

【選択図】 図 1

認定・付加情報

特許出願の番号	特願2001-202918
受付番号	50100975534
書類名	特許願
担当官	第七担当上席 0096
作成日	平成13年 7月 5日

<認定情報・付加情報>

【提出日】	平成13年 7月 4日
-------	-------------

出 願 人 履 歴 情 報

識別番号 [000005108]

1. 変更年月日 1990年 8月31日

[変更理由] 新規登録

住 所 東京都千代田区神田駿河台4丁目6番地

氏 名 株式会社日立製作所